

USGS Instructional Memos:
A brief summary

ESIP Data Stewardship Monthly Meeting
June 15, 2015

Viv Hutchison

Why Manage Our USGS Data?

- Data are a lasting resource that the USGS collects and uses.
- The USGS cannot operate efficiently without good, solid, accurate, reliable, useful, and timely data. Data are expensive corporate assets with value beyond our immediate need or lifetime.
- Data Management:
 - creates efficiencies in how science is done.
 - Improves provenance and reproducibility in the science iteration process.
 - Supports scientific review and integrity
 - Maximizes the effective use and value of data assets
 - Allows data to be more discoverable, understood, and enables better data integration.



Open Data Initiatives

- Executive Order -- Increasing Public Access to the Results of Federally Funded Scientific Research. Feb. 22, 2013
- Executive Order -- Making Open and Machine Readable the New Default for Government Information May 9, 2013
- M-13-13-Open Data Policy-Managing Information as an Asset

USGS Instructional Memos*

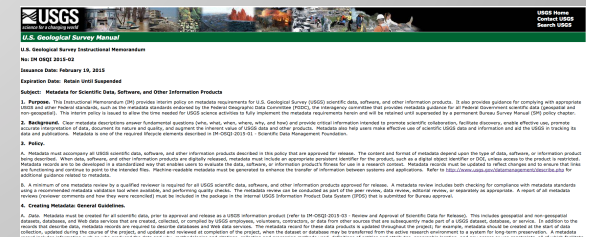
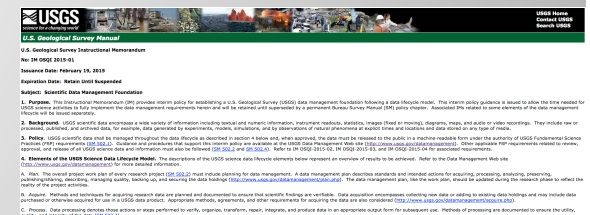
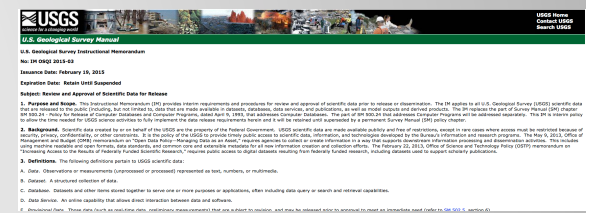
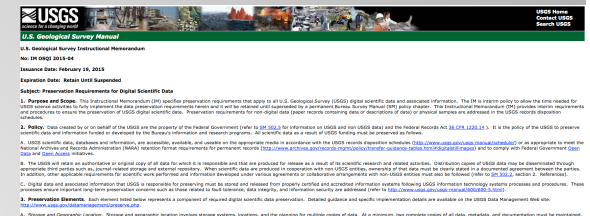
- Scientific Data Management Foundation
- Metadata for Scientific Data, Software, and Other Information Products

- Review and Approval of Scientific Data for Release

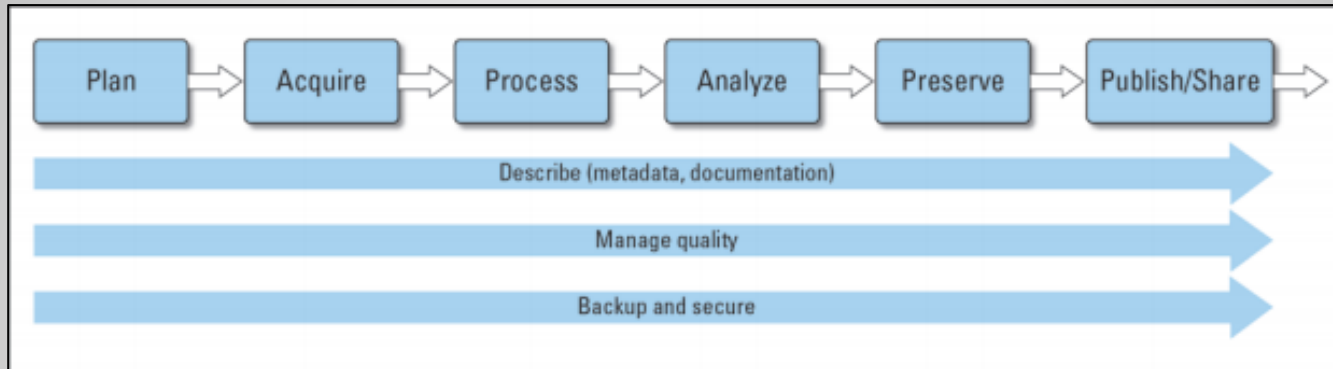
- Preservation Requirements for Digital Scientific Data

- <http://www.usgs.gov/usgs-manual/95imlist.html>
 - Near the bottom of the webpage

*Released as IMs; will become policy in October 2016. Developed by a small team of people “Data Policy Team” through the Community for Data Integration (CDI)



Scientific Data Management Foundation



- USGS scientific data shall be managed throughout the science data lifecycle.
- USGS scientific data must be released to the public in a machine-readable form. (such as XML, JSON, CSV)
- A Data Management Plan (DMP) must be included as part of or an addendum to the Project Work Plan for every project.
- Center Directors are responsible for ensuring these IMs are met.

Metadata for Scientific Data, Software, and Other Information Products

- All data* approved for release must be accompanied by standardized metadata
- Metadata for data must:
 - using FGDC or ISO suite of standards
 - be robust, meaningful, and complete
 - contain a digital object identifier (DOI)
 - be updated to reflect changes and ensure links are viable
 - be reviewed (1 reviewer) for quality of content and standards compliance. Comments are placed in IPDS
 - be shared through the Science Data Catalog
- Metadata for *software and other information products* is also required (it is different from metadata for data)

*data = non-geospatial AND geospatial data, databases, and web services

Southwest Non-native Invasive Plant Database (SWEMP07)

Metadata:

- Identification Information
- Data Quality Information
- Spatial Data Organization Information
- Spatial Reference Information
- Policy and Access Information
- Distribution Information
- Metadata Reference Information

Identification Information:

Citation:

Citation Information:
Originator:
Kathryn Thomas and Patricia Guenther, U.S. Geological Survey, Southwest Biological Science Center (USGS-SWSC)
Publication Date: 20070508
Title: Southwest Non-native Invasive Plant Database (SWEMP07)
Geospatial Data Presentation Form: tabular digital data
Online Linkage:
<http://bioc.usgs.gov/research/projects/swemp/swemp07.asp>

Description:

Abstract:
The Southwest Exotic Plant Mapping Program (SWEMP) is a collaborative effort between the United States Geological Survey and federal, tribal, state, county and NCO partners in the southwest. This project is an ongoing effort to compile and distribute regional data on the occurrence of non-native invasive plants in the southwestern United States. The database represents the known area (represented by a point location, i.e. site) of non-native invasive plant infestations within Arizona and New Mexico, and adjacent portions of California, Colorado, Nevada and Utah. These data, collected from 1911 to 2006, represent the field observations of various state, federal, tribal and county agencies, along with some specimen data from herbaria. The SWEMP database is published at least once a year and consists of a compilation of all data submitted up to the date of publication.

Purpose:
This dataset was created to provide a regional perspective on non-native invasive plant distributions. It can be used to assist land managers, as well as the public, to review the locations and extent of reported infestations. These data can ultimately help guide management strategies and policies for the control of non-native invasive plant species. All plant species in the database are non-native as defined by the USDA PLANTS database (2007), the extent to which they are invasive has not been determined.

Time Period of Content:

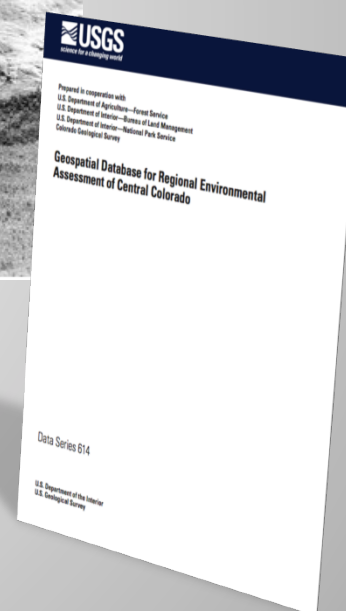
Time Period Information:
Range of Dates/Times:



Review and Approval of Scientific Data for Release

Open Data Policies:

For the first time for every scholarly publication (*USGS series reports and journal articles, etc.*), the data on which the scholarly conclusion is based must be released at the same time as the scholarly publication.



Review and Approval of Scientific Data for Release

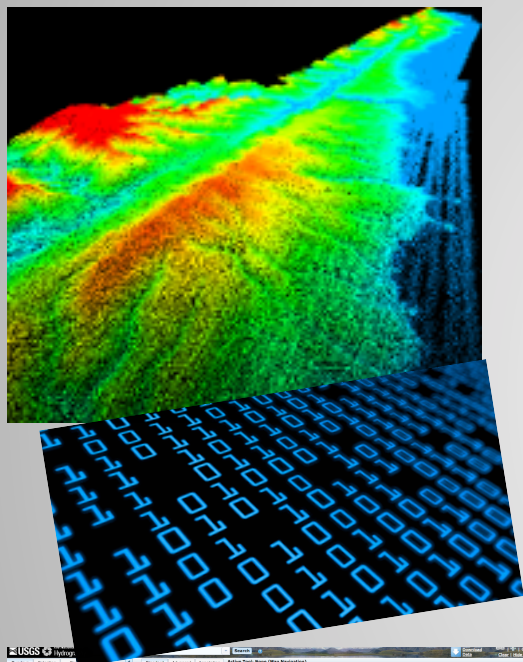
- USGS releases both approved and provisional data
 - provisional data must contain a disclaimer until formally released.
- Approved data must:
 - be assigned a persistent identifier, specifically a Digital Object Identifier (DOI)
 - be accompanied by a citation
 - be accompanied by FGDC or ISO standard metadata



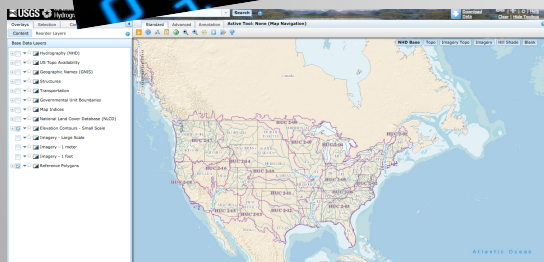
John A. Engott, 2015, Mean annual water-budget components for the Island of Oahu, Hawaii, for average climate conditions, 1978-2007 rainfall and 2010 land cover: U.S. Geological Survey Data Release, <http://dx.doi.org/10.5066/F7XP72ZX>.



Review and Approval of Scientific Data for Release



- Two reviews: 1 each for data and metadata
 - Different from peer review for publications
- Data must be managed and distributed through a data system that can ensure the long-term preservation, discoverability, accessibility, and usability of the resource
- Most data approved for release is approved by the Science Center Director or their designee through IPDS
 - interpretive data must be approved by a Bureau Approving Official

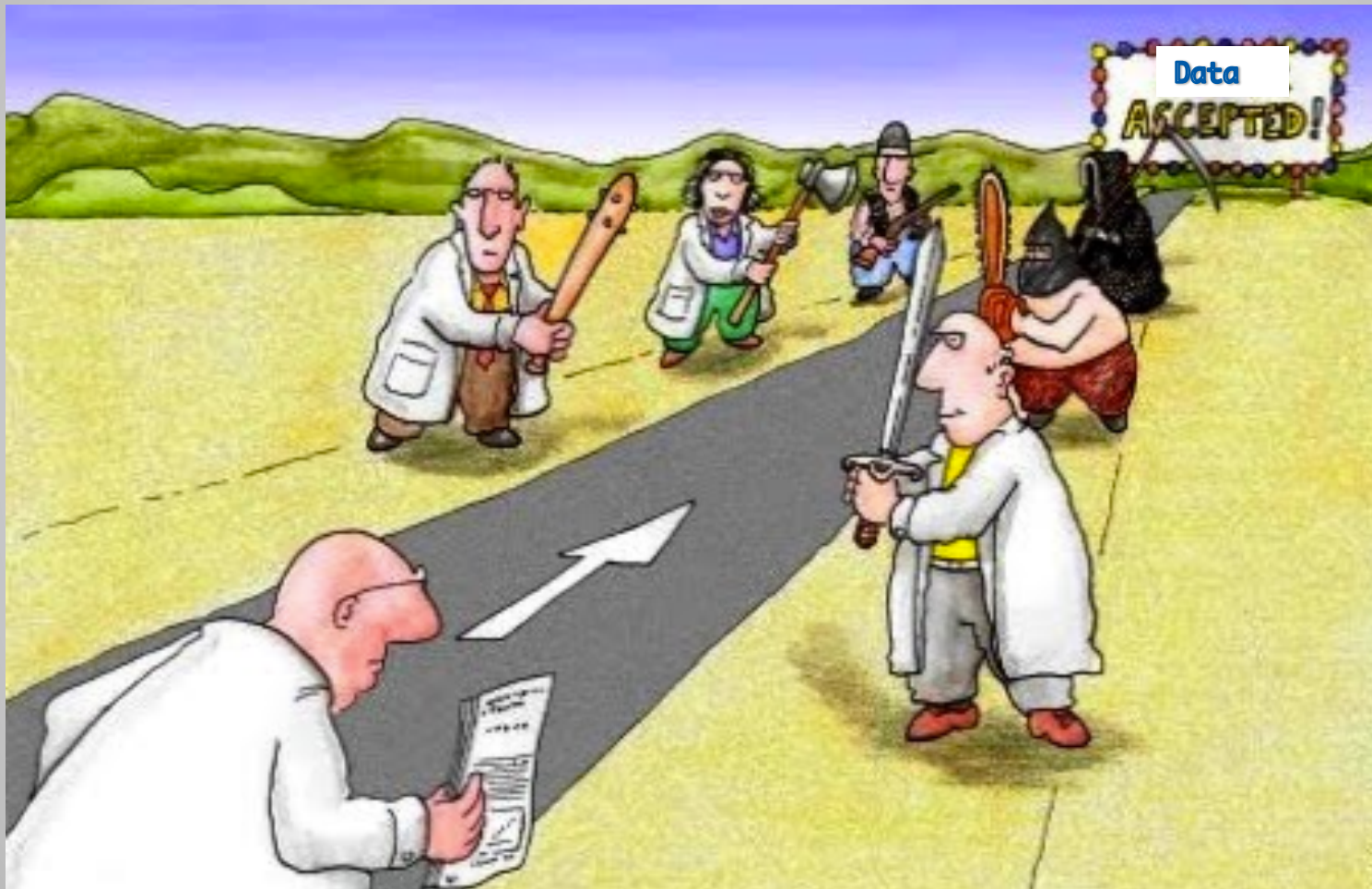


Preservation Requirements for Digital Scientific Data

- All scientific data produced as a result of USGS funding must be:
 - preserved
 - accessible
 - available
 - useable on the appropriate media
 - stored and released from properly certified and accredited information systems
- The USGS will retain an authoritative or original copy of all data for which it is responsible
- Elements of data preservation:
 - Storage and geographic location
 - Data viability and integrity
 - Information security
 - Metadata
 - File Formats



Sometimes the road to data release seems
a bit daunting...



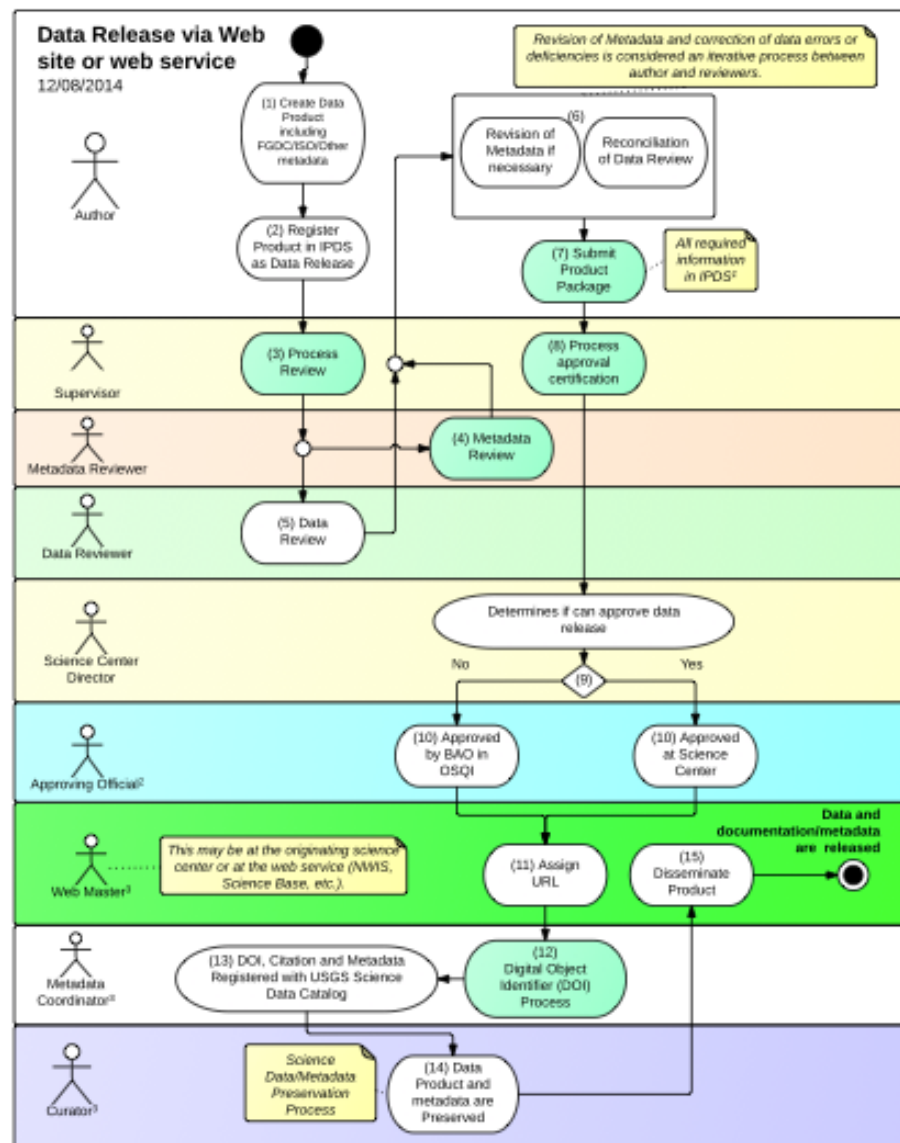
So, we are trying to ease the journey...

Workflow Example for Data Release

Scientist creates a data product to disseminate via a USGS Web page or USGS Web service.

Data release involves:

- review of metadata and data content
- use of the USGS Information Product Data System (IPDS);
- acquisition of Digital Object Identifier (DOI) for data;
- approval by the designated approving official. A Science Center Director (or designee) can approve non- interpretive data. A Bureau Approving Official (BAO) must approve data that are considered new interpretive;
- release through USGS Web site or USGS-maintained web service.



1. IPDS is the Bureau Information Product Data System. Refer to URL: <http://internal.usgs.gov/publishing/ipds.html>
 2. Under FSP, approval of non-interpretive information products is delegated to the Science Center Director
 3. The Web Master, Metadata Coordinator, and Curator are data management roles and may be delegated to the same person or different individuals at the discretion of the Science Center Director or if a web service like NWIS or USGS Science Base, will be handled by the service.

Tools for Aiding Data Release

Data Management Website (www.usgs.gov/datamanagement)

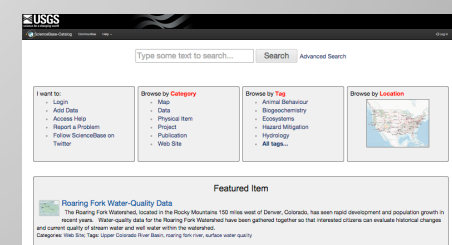
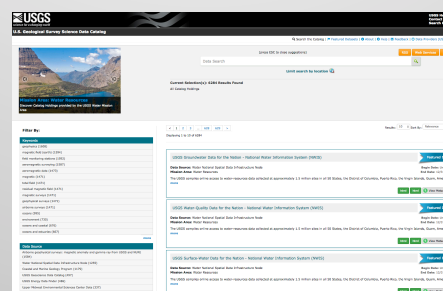
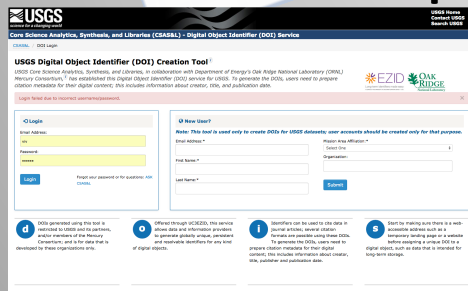
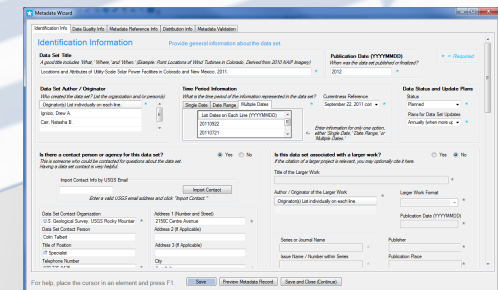
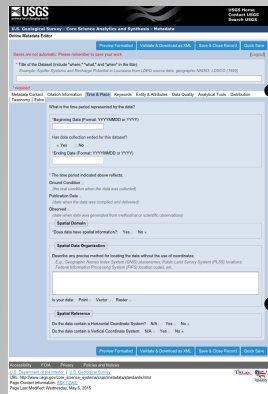
Data Management Plans:

- DMP Tool template
- Tools in ScienceBase

Documentation Tools:

- **Metadata Wizard**
- **Online Metadata Editor**
- **DOI Tool**

- **Data Release Option: ScienceBase**
- **Science Data Catalog**
- **Frequently Asked Questions (posted to USGS website as completed)**



USGS Community for Data Integration

- community of practice
- open to all
- focused on:
 - advancement of scientific data & information management
 - data integration capabilities across the USGS and external organizations to enhance earth science research
- encourages innovative ideas across the community through:
 - working groups
 - seed funding for projects – RFP announced in Sept/Oct timeframe generally
- <http://www.usgs.gov/cdi/>

